

## ФОРМАЛЬНАЯ МАТЕМАТИЧЕСКАЯ МОДЕЛЬ СИНТЕЗА МАССИВА БИЛЛИНГОВОЙ ИНФОРМАЦИИ

*Аннотация.* В статье рассмотрен способ создания формальной математической модели, применяемой для синтеза биллинговой информации в целях получения массивов, пригодных для тестирования корректности реализации поисково-аналитических алгоритмов.

*Ключевые слова:* биллинг; математическая модель; генерация.

С целью синтеза массивов биллинговой информации для тестирования информационно-аналитических комплексов предложена пространственно-временная событийная модель жителей населенного пункта в терминах телекоммуникаций [1, 2]. В качестве основы модели выбрана популяция абонентов операторов сетей сотовой связи, проживающих на вымышленной территории, перемещающихся по заданным правилам в течение генерируемого периода времени и осуществляющих взаимодействие путем совершения звонков и отправки SMS.

Модель учитывает статистические распределения биллинговой информации и социальные характеристики абонентов операторов сотовой связи (круг общения: наличие родственных, личных связей). Социальные характеристики представлены шаблонами поведения для различных социальных групп, включая характеристики взаимодействия в сетях сотовой связи и характеристики перемещения людей в рамках населенного пункта.

Модель синтеза биллинговой информации основывается на поведении абонентов с точки зрения сети операторов связи, является совокупностью модели перемещений абонентов в течение заданного промежутка времени  $MSH$  и модели соединений  $MS$ .

Модель перемещений  $MSH$  имитирует перемещения абонентов в заданном временном интервале в рамках населенного пункта, который представляется квадратом, для каждой клетки которого задаются модифицируемые списками параметры LAC и CellID базовых станций.

Перемещения абонентов могут моделироваться в двух совмещаемых режимах: на основе частотных распределений по популярности конкретных зон или на основе шаблонов перемещений. В первом случае перемещения абонентов

имеют большую случайность, возможность быстрого редактирования характера перемещений абонентов путем изменения приоритетов конкретных зон (ячеек карты), во втором — перемещения предсказуемы и направляемы (пользователь может менять маршруты абонентов и указывать конкретные точки маршрута по своему усмотрению).

В состав модели вводятся абоненты, каждый из которых имеет пол, возраст, может иметь семью:

- множество абонентов оператора сотовой связи  $H$ :  $H = \{h_i\}_{i=1}^{n_h}$ , где  $n_h$  — количество абонентов;
- множество «семей»  $W$ :  $W = \{w_i \mid w_i \subseteq H\}_{i=1}^{n_w}$ , где  $n_w$  — количество.

Абонентские терминалы представлены в модели множествами номеров IMSI, IMEI и MSISDN:  $N_{\text{imsi}}$ ,  $N_{\text{imei}}$ ,  $N_{\text{msisdn}}$  с количеством элементов, равным  $n_h$ :

$N_{\text{imsi}} = \{n_{\text{imsi}_i}\}_{i=1}^{n_h}$ ,  $N_{\text{imei}} = \{n_{\text{imei}_i}\}_{i=1}^{n_h}$ ,  $N_{\text{msisdn}} = \{n_{\text{msisdn}_i}\}_{i=1}^{n_h}$ . Для формирования элементов множеств используется алгоритм генерации неповторяющихся чисел требуемого формата.

Совокупность абонентов и соответствующих им абонентских терминалов формирует абонентскую базу (базу принадлежности).

Передвижение абонентов в рамках населенного пункта описывается шаблонами перемещений  $SH$ . Количество шаблонов произвольное, задается значением  $n_{sh}$ ,  $SH = \{sh\}_{i=1}^{n_{sh}}$ . Шаблон перемещений представляет собой упорядоченный по времени список LAC и CellID, по которому отслеживается передвижение абонента в течение заданного времени.

Населенный пункт характеризуется массивом ячеек его карты: двумерным вектором  $\text{map}$ ,  $n_{\text{map}}$  — его длина (количество ячеек карты), в каждом из элементов  $\text{map} \in (1, n_{\text{map}})$  которого содержатся два параметра базовой станции, обслуживающей участок территории (LAC и CellID).

При генерации массива абонентов  $H$  каждому абоненту  $h_i$  присваивается номер, соответствующий шаблону  $sh_i$  из массива шаблонов  $SH$ . Соответственно, каждому абоненту назначается начальная точка на карте и перечень точек его местонахождения в последующие моменты времени.

В модели соединений  $MS$  формируется массив строк биллинга  $D$ :  $D = \{d\}_{i=1}^{n_d}$ ,  $n_d$  — количество строк. Каждая строка соответствует событию (действию) в сети. Строки массива  $D$  формируются из массивов, описывающих события различного типа:  $T$ ,  $G$ ,  $C$  и  $S$ :

- $T$  — массив перемещений:  $T = \{t\}_{i=1}^{n_t}$ , где  $n_t$  — количество;
- $G$  — массив подключений к GPRS:  $G = \{g\}_{i=1}^{n_g}$ , где  $n_g$  — количество;
- $C$  — массив звонков:  $C = \{c\}_{i=1}^{n_c}$ , где  $n_c$  — количество. Массив звонков  $C$  состоит из массивов исходящих  $C_o$  и входящих  $C_i$  звонков:  $C = C_i + C_o$ , причем  $n_c = n_{ci} + n_{co}$ ,  $n_{oi}$  и  $n_{ci}$  — количество.

- $S$  — массив СМС:  $S = \{s\}_{i=1}^{n_s}$ , где  $n_s$  — их количество.  $S$  состоит из массивов исходящих  $S_o$  и входящих  $S_i$  СМС:  $S = S_i + S_o$ , причем  $n_s = n_{si} + n_{so}$ ,  $n_{si}$  и  $n_{so}$  — количество.

Массив соединений  $Y$  является совокупностью массива звонков и массива СМС:  $Y = C + S$ , где  $n_y$  — количество соединений ( $n_y = n_c + n_s$ ). Характеристики элементов массива соединений:

а) идентификатор типа соединения между абонентами  $A \in (\overline{1, n_a})$ , где  $n_a$  — количество различных типов соединений;

б) идентификаторы абонентов, участвующих в соединении:

– абонент  $h_1 \in H$  — источник соединения (ИС);

– абонент  $h_2 \in H$  — получатель соединения (ПС);

в) номера телефонов абонента А (ИС) и Б (ПС), где  $n_p$  — предельное значение данного параметра: ИС  $p_1 \in (\overline{1, n_p})$ , ПС  $p_2 \in (\overline{1, n_2})$ .

При формировании массива соединений  $Y$  учитываются статистические характеристики биллинговой информации:

а)  $K_0 = \langle F_h, F_f, F_{\text{time}}, F_{\text{dur}}, F_n, F_a \rangle$  — не связанные с адресацией соединения, описываемые функциями распределения:  $F_h$  — вероятности генерации типа события,  $F_f$  — длительности события,  $F_{\text{dur}}$  — продолжительности,  $F_{\text{time}}$  — времени,  $F_a$  — типа,  $F_n$  — количества соединений;

б)  $K_1 = \langle F_l, F_{t1} \rangle$  и  $K_2 = \langle h1, F_{t2} \rangle$  — связанные с адресацией соединения:  $K_1$  — от ИС к ПС,  $K_2$  — от ПС к ИС.  $F_l$  — функция распределения выбора ПС, а  $F_{t1}$  и  $F_{t2}$  — функции распределения промежутков времени между началами инициализации двух последовательных соединений для ИС и ПС.

Таким образом, структура соединений определена вектором  $\langle Y_i \rangle_{i=1}^{n_y}$  и описывается выражением:  $Y_i = \langle A, F_{h1}, F_{h2}, p_1, p_2, K_0, K_1, K_2 \rangle$ .

Модель соединений  $MS$  определяется выражением:

$$MS = \langle H, W, \langle Y_i \rangle_{i=1}^{n_y}, K_0 \rangle.$$

С учетом перемещений абонентов по заданным шаблонам модель синтеза биллинговой информации описывается выражением:

$$M = \langle H, W, \langle Y_i \rangle_{i=1}^{n_y}, F_{\text{time}}, F_{\text{dur}}, F_a, F_n, MSH \rangle.$$

Разработанная пространственно-временная событийная статистическая модель позволяет создать на ее основе программный комплекс, решающий задачу синтеза массивов биллинговой информации.

### Список литературы

1. Семенищев И. А., Синадский А. Н., Синадский Н. И. Алгоритм формирования массива биллинговой информации на основе статистической модели поведения абонентов сотовой связи // XV Всерос. науч.-практ. конф. студентов, аспирантов и молодых ученых : сб. материалов. Курган : Курган. ГУ, 2016. С. 199–203.
2. Семенищев И. А., Синадский А. Н., Синадский Н. И. Статистические характеристики массива биллинговой информации при моделировании поведения абонентов сетей сотовой связи // 12-я Международ. молодеж. науч.-тех. конф. : сб. материалов / под ред. А. А. Савочкина. Севастополь : Севастопол. гос. ун-т, 2016. С. 207.

УДК 004.056

**П. В. Сушков**

Научный руководитель: канд. тех. наук, доц. Н. И. Синадский  
Уральский федеральный университет, Екатеринбург

### **МОДИФИКАЦИЯ АЛГОРИТМА БАРАБАШИ — АЛЬБЕРТ ПРИ СИНТЕЗЕ МАССИВОВ ДАННЫХ, ИМИТИРУЮЩИХ ВЗАИМОДЕЙСТВИЕ В СОЦИАЛЬНЫХ СЕТЯХ**

*Аннотация.* Рассмотрен метод синтеза массива данных, имитирующих взаимодействие пользователей в социальных сетях, на основе модифицированного алгоритма Барабаши — Альберт с учетом заданного шаблона взаимодействия пользователей.

*Ключевые слова:* модель Барабаши — Альберт; имитация взаимодействия пользователей социальных сетей.

Изучение программных средств анализа больших данных в рамках реагирования на компьютерные инциденты, предусмотренное по дисциплинам специальности «Информационная безопасность телекоммуникационных систем», требует наличия учебных массивов информации о взаимодействии пользователей в социальных сетях [1–3].

С целью синтеза таких массивов предложен алгоритм, базирующийся на известной модели Барабаши — Альберт, позволяющей синтезировать случайные безмасштабные сети с использованием принципа предпочтительного присоединения [4]. Принцип предпочтительного присоединения заключается в том, что в каждый момент времени в синтезируемую сеть добавляется новый узел,